心理科学进展 2023, Vol. 31, No. 2, 1–13
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.001

基于大数据的文化心理分析*

吴胜涛[1] 茅云云[1] 吴舒涵[2] 冯健仁[3]
张庆鹏[3] 谢 天[4] 陈 浩[5,6] 朱廷劭[7]

([1]厦门大学社会与人类学院; [2]厦门大学新闻与传播学院, 厦门 361005)
([3]广州大学公共管理学院, 广州 510006) ([4]武汉大学哲学学院, 武汉 430072)
([5]南开大学周恩来政府管理学院, 天津 300350) ([6]中山大学广州粤港澳社会心理建设研究中心,
广州 510006) ([7]中国科学院心理研究所, 北京 100101)

摘 要 随着大数据技术与文化心理学的融合发展, 计算文化心理学作为一门新兴交叉学科逐渐兴起, 大尺度、近乎全样本的文化心理分析真正得以实现。计算文化心理学关注的文化变量主要围绕个人主义/集体主义这一文化心理学使用最为广泛的维度展开, 分析方法包括特征词典、机器学习、社会网络分析、仿真模拟等, 分析思路包括时间维度上的文化变迁效应以及空间维度上的文化地理效应。当然, 计算文化心理学在为传统文化心理研究提供新方法、新范式的同时, 也存在解码失真、样本偏差、词语多义性、隐私风险等局限, 未来研究应重视变量理论解释、文化动态演化分析、学科深度整合、生态效度等问题。

关键词 大数据, 文化, 计算文化心理学, 文化变迁, 文化地理
分类号 B949; C91

早期大数据心理研究多集中在基本情绪、消费决策等相对直观的心理概念上, 针对文化心理等抽象心理概念的大数据研究则相对较少(Golder & Macy, 2011, 2014)。进而, 随着大数据技术与文化心理学的融合发展, 采用大数据思维、大数据资源和计算工具来理解文化符号及其互动过程的计算文化心理学也快速发展起来, 这使得大尺度甚至近乎全样本的文化心理分析真正得以实现。本研究回顾了计算文化心理学的产生过程、变量提取和计算方法, 以及使用大数据技术在时间维度上分析文化变迁效应、在空间维度上分析文化地理效应的思路、方法和主要发现。最后, 本研究反思了计算文化心理学的局限, 并对其未来发展方向进行了展望。

1 大数据时代的文化心理学

大数据是一种容量大、种类多、速度快的信息资源, 需要特定计算技术和分析方法挖掘其中的价值(de Mauro et al., 2015, 2016)。由于大数据巨大的潜在价值, 近年来大数据分析技术迅速在各个研究领域中展露头角, 逐渐成为社会科学的新工具(陈浩 等, 2013; Kosinski et al., 2015)。尽管大数据分析技术的历史并不长, 但从近 10 年的大量文献来看, 它已经成为当前哲学、自然科学和社会科学共同关注的重要问题, 为人文与社会科学提供了继实验、理论、仿真之后的第四种研究范式, 并催生了计算社会科学的诞生(Gray, 2009; Lazer et al., 2009)。

随着大数据时代的到来, 传统文化心理学面临诸多挑战, 但也迎来了范式变革的机遇。首先, 互联网将人与人之间的距离拉近, 使得不同文化背景的人主动或被动地汇聚在一个时空压缩的“地球村”里, 全球范围的文化互动、变迁成为常态, 但由于研究资料受限、分析方法滞后, 传统文化心理学在进行大尺度时间、空间分析时费时费力。

收稿日期: 2022-01-14
* 国家社会科学基金重点项目(17AZD041)及厦门大学校长基金(20720221017)的支持。
通信作者: 吴胜涛, Email: michaelstwu@xmu.edu.cn or wust2011@163.com

幸运的是, 信息技术的发展使得不同时间、空间的文化符号及其互动过程留下了海量的数据印记, 这为研究者分析不同时间、不同空间下的文化心理特征提供了无可比拟的大规模生态数据库。进而, 大数据技术赋能研究者们方便地抓取网络生态数据, 研究文化动态演化及节点事件效应(Park, Baek et al., 2014; Wu, Zhou et al., 2018), 使得时间趋势和空间地理上的大尺度分析真正成为可能, 文化心理学的研究问题也得以拓展。

其次, 大数据时代文化符号生产快速增长, 但是基于纸笔测验和行为实验的传统文化心理学研究往往是滞后的, 且样本量较小、代表性不够, 远远不能满足研究者和决策者的需要。而基于大数据的文化心理学研究能够以相对较低的时间、人力成本实现近乎全样本的数据进行自动抓取、实时分析乃至实验操控, 并从海量数据中迅速锁定、预测某一群体的行为特征, 进而实现宏观群体水平的文化心理分析、预测和改变。

再者, 大数据时代的文化互动日趋个体化和复杂化, 而传统的文化心理学容易忽视文化成员在个体水平及复杂情景中的差异, 仅通过抽样追求简单因果关系, 却难以把握文化心理、行为间的复杂相关性。也就是说, 在没有搞清楚"是什么"的情况下就去分析"为什么" (顾肃, 2021)。巨大、连续的数据存储和模型分析, 以一种细粒度和大规模乃至全球范围的方式实时研究个体行为, 能够根据个体和情景变异进行适应性的调整和持续、实时的纠正; 同时, 庞大数据量补偿了精确性上的损失, 这有助于分辨出最合适的相关关系, 捕捉小数据测量时没有的新价值(Lazer et al., 2009)。尽管基于大数据技术的词频、预测模型或社会网络分析并非新鲜事物, 但将其应用于文化心理学研究能够从整体上把握对象, 促进从因果思维向相关性思维转变, 进而有利于全面、客观地揭示隐藏在数据背后的文化意涵和规律(陈云松, 2015)。

最后, 更为重要的是, 大数据时代的文化形态也发生了深刻变化。随着互联网的快速发展与网民的大量增加, 基于网络的独特文化现象层出不穷, 一些经典的文化行为特征在网络上的表现也有所不同, 但传统文化心理学对此知之甚少。例如, 研究发现相对于其他社会环境, 网络环境下人们的情绪表达会更极端, 而且网络使用的时间越长, 极端情绪越明显(King, 2001)。此外, 由于线上与线下任务的认知加工过程不同, 人们在一些经典的文化认知表现上也存在差异。例如, 相对于认知负荷较大的线下记忆任务, 低认知负荷的线上浏览任务中被试有更多的认知资源来处理外部信息, 因此在社会行为推理上会表现出较小的对应偏差, 基于国别的文化差异被基于线上和线下的文化差异所代替(Miyamoto & Kitayama, 2002)。尤其值得注意的是, 移动互联网本身就是一种带有偏差的文化产品, 例如具有相应文化倾向(如高个人主义、低不确定性回避)的用户也更能体验到技术−文化匹配后的愉悦感和效用感(Lee et al., 2007)。因此我们不能仅满足于将研究线下文化现象的思维来简单套用到线上行为的文化心理研究中。

总之, 随着计算机和大数据技术的深入发展, 人类社会及其文化形态正经历着深刻的变革。文化符号的生产和互动日趋复杂, 文化成员及其社会网络更是留下了海量的文本与行为足迹。这使得文化必须被快速地描述、预测甚至改变, "可计算"的文化符号及其互动过程也逐渐成为文化心理学的研究对象。 接下来, 我们将主要对计算文化心理学的产生过程、主要变量和分析技术进行回顾, 并介绍该技术在相关文化变迁、文化地理效应的应用。

2 计算文化心理学的兴起

2.1 计算文化心理学的产生

在计算文化心理学正式产生以前, 文化学者就受到计算机技术的启发, 提出了"文化计算"概念, 强调要发挥计算机的独特优势, 以便多维度展现文化及其与人的互动(Tosa et al., 2005)。通过文化计算, 研究者首先在艺术领域实现了文化的数据化呈现和翻译(Tosa et al., 2019)。进而, 文化计算把计算及相关技术应用到越来越多的文化领域, 探索其发展规律、提示其内在联系并对其进行量化分析与展示的科学(赵海英 等, 2016)。如Michel 等人(2011)通过对谷歌图书项目的数据进行分析, 发现了特定关键词使用频率的时间变化规律, 并由此推导出人类文化发展的趋势, 如语法演变、集体记忆和流行病学的变化趋势等; 吴育锋等人(2018)通过"文心" (中文心理分析系统)对小说《平凡的世界》中的对话文本进行文学智

能分析，发现其计算出的小说人物性格与以往文艺研究结果一致，并发现了不同时代人物呈现出性格变迁的趋势(如新时代青年的性格更加开放、外向)；陈善雄等人(2019)基于深度学习的方法进行古彝文识别，通过计算机再现中华民族传统文化；邢付贵和朱廷劭(2021)采用 N-Gram、互信息、信息熵、位置成词概率等多特征融合的新词发现技术，在文言文语料库里抽取古汉语词汇，形成古文词典，并在此基础上实现对文言文的有效分词，发现《四库全书》中儒家经典更加注重道德实践，而非伦理思辨(Hu et al., 2021)，河南历史名人对于事物之间差异的认知在春秋战国时期最高，到汉朝开始降低(Zhao et al., 2021)。当然，文化计算不仅局限于文本资料，如 Borth 等人(2013)基于图像和视频等视觉内容的情绪计算方法，建立了大规模视觉情感本体和视觉概念检测库，显著提高了情绪检测准确率。

Neuman 等人(2012)在“文化计算”的基础上正式提出了“计算文化心理学”的概念，即通过计算工具和大数据资料库来研究文化符号及其互动过程的心理学分支。值得注意的是，虽然文化计算和基于计算的文化心理学有很多相似之处，但是二者也有区别：前者侧重文化载体的呈现，以及文化基因的提取；后者除了一般意义上的文化计算，还注重分析人与文化环境的交互关系，并且文化载体不仅是传统艺术文化，也包含网络社交媒体等。下一节，我们将对计算文化心理学的研究内容，即核心变量的提取和计算方法，进行系统介绍。

2.2 文化心理变量的提取

为了理解文化现象，心理学者需要首先提取文化心理变量，进而理解影响其成员思维的符号系统和互动机制。语言是被分享的文化系统(王斌, 2012)，因此我们可以把语言看做是理解文化的窗口以及提取文化心理变量的来源(罗伯特·怀尔等, 2017)。语言的本质是思想、话语的直接现实和展示(吴美川, 2020)，其在计算文化心理学领域的语言载体主要包括书籍、社会媒体数据、档案数据、网络音频等。相对而言，基于互联网的社交媒体在学术研究中被使用的频率更高。一方面，社交媒体用户的样本量大甚至接近总体，且社交媒体数据便于研究者采用自下而上的数据驱动方式验证已有的文化心理理论，这比理论驱动实施起来更容易；另一方面，社交媒体带来了新的文化心理现象，是值得探索的新领域(喻丰 等, 2015; Qiu et al., 2018)。

针对语言文化载体，研究者主要提取的文化心理变量包括个人主义/集体主义、个性解放/自我表达、文化松紧度等，详见表 1。这些文化心理变量之所以被研究者们重点关注，是因为：第一，这些变量本身具有重要研究价值，如个人主义/集体主义是文化心理学中使用最为广泛的文化维度，尤其对社会文化变迁和跨文化比较研究具有重要意义；第二，社会现实的需求，如新冠疫情的全球大流行显著增加了规范遵从的强度，这需要研究者及时分析其文化适应后果及政策干预策略；第三，此类文化变量的可操作性强，便于用大数据方法来进行抓取和计算。

此外，计算文化心理学研究也对一些文化变量进行了操作化和创造性的发展。首先，考虑到大数据的语料库特点，选用一些便于操作的语言特征作为经典文化心理学变量的指标，例如用第一人称单数的使用频率作为个人主义的指标(Twenge et al., 2013)；其次，将个体的行为印记(如@他人、成组人群比例)作为文化心理变量的指标，是对以往基于主观报告的文化心理测量的重要发展(Wu, Li et al., 2018)；此外，根据新媒体的特点，采用表情包等作为文化相关情绪表达规范的指标也是一种有益的创新(Koda & Ishida, 2006)。

2.3 文化心理变量的计算方法

大数据时代下，文化心理学有机会克服以往通过统计抽样来代替总体样本的取样难题，研究结论能够发现普遍规律但无法将其用之于个体的窘境，以及必须在实验室中创设各种非真实情境的生态效度问题(喻丰 等, 2015)。因为大数据通过网络爬虫等技术，可以获得几乎接近总体的样本，并且这些几乎是自然条件而非控制条件下产生的。同时，从这些数量大且自然化的数据中，个体作为参与主体的特性得以突出，这使得研究者在得到普遍规律时可以不放弃个体的个性规律。

目前采用大数据技术进行文化心理分析主要有 4 种方法：第一，基于特征词典的频次分析，如通过对数字、表情和语言等方面的关键特征分析，研究个体或群体的文化价值观，这是一种简单有效但相对粗糙的分析(李国杰, 程学旗, 2012)。第二，

表 1 计算文化心理学的常用文化心理变量、计算方法和研究举例

文化心理变量/研究问题	计算方法	操作过程	主要发现
个人主义/集体主义的变迁	基于特征词典的频次分析	个人主义为第一人称单数(如 I, me, my, mine, myself)或行为特征词使用频率(如 get, choose); 集体主义为第一人称复数(如 we, us, our, ours, ourselves)或行为特征词使用频率(如 give, obliged) (Greenfield, 2013; Twenge et al., 2013; Yu et al., 2016; Zeng & Greenfield, 2015)	Google 电子图书语料库显示, 过去 200 年、特别是二战以后, 世界范围内出现了个人主义文化增强、集体主义文化式微(或稳定)的历史变化趋势。
	社会网络分析	集体主义为微博用户@他人的次数(Wu, Li et al., 2018)。个人主义是独自行走者的比例, 集体主义是成群行走者的比例(Favaretto et al., 2016)。	2010~2016 年间, 短暂的经济增长放缓伴随集体主义的显著增长。集体主义文化(如中国)下 成群行走者比例高于个人主义文化(如澳大利亚)。
个人主义、文化松紧度与情绪表达规范	基于特征词典的频次分析	基于表情符号的 Unicode 标注, 把 emoji 表情转译成文本, 然后再用 LIWC 对每个表情符号进行积极情绪、消极情绪计算和排序(Lu et al., 2016)。情感表达规范为 Facebook 用户状态更新中积极(或消极)情绪词汇的使用频率, 文化松紧度的分数是根据学校体罚的正当性、违反法律的处罚程度等国家统计的 9 项综合指标计算得出的(Liu et al., 2018)。	个人主义国家或地区的用户更倾向于使用正面的表情符号, 而较少使用负面的表情符号; 来自松文化地区的人更可能表达积极的情绪, 而不太可能表达消极的情绪。
个人主义与个性解放/自我表达的地区差异	机器学习/人格预测模型	基于 Twitter 个人主义指标(第一人称单数使用率)与已有国家个性解放/自我表达得分(自治、平等、选择与言论自由)建立回归模型, 然后用其他国家地区的个人主义指标预测其个性解放/自我表达得分(吴胜涛 等, 2018)。	个人主义正向预测自治、平等, 即: 个人主义越强的国家或地区, 越倾向于自治、平等的价值观。
性别/种族平等、个人主义生活/工作态度的变迁	机器学习/词嵌入联想测验	基于 Google News & Books 数据集的词向量或词嵌入文件, 以靶词和属性词的语义关联(余弦相似度)作为态度指标, 分析其时间变化趋势 (Garg et al., 2018; Hamamura et al., 2021)。	1910~1990 年间, 性别偏见和种族偏见随着时间的推移而减少。1950~1990 年间, 个人主义/集体主义与工作、成就词的关联下降。
不确定规避、个人主义与社交距离和创新扩散	社会仿真模拟	社交网络距离为连接每两个主体之间最短路径的平均值; 创新扩散为由外部随机决定的 t = 0 处的特定位置(Desmarchelier & Fang, 2016)。	不确定规避对创新扩散速率有负向影响, 而个人主义对扩散速率有正向影响。

基于机器学习的模型预测或词嵌入联想测验, 如通过机器学习模型对用户的个性、价值观和意识形态进行识别, 这种方法更精确, 对心理特征的把握也更全面(Kosinski et al., 2015); 词嵌入联想测验以词嵌入向量特征代替传统的词频特征, 将靶词和属性词向量的余弦相似度作为语义关联度的指标, 是社会态度与价值观研究的新方法(Caliskan et al., 2017; Hamamura et al., 2021)。第三, 社会网络分析, 如通过对网络搜索、网友互粉等数据的分析, 研究不同文化群体的行为和决策风格, 这种方法目前在文化心理学领域还相对较少, 但具有重要的社会实践价值(Wu, Li et al., 2018); 第四种, 社会仿真模型, 因为很多社会事件是无法在现实中进行实验的, 所以采用仿真模拟的办法来研究某一特定的系统和策略, 从而达到分析社会事件的目的(刘婷婷 等, 2016)。当然,

这 4 种技术并不是孤立的, 研究者也会根据研究问题采取一种或多种技术。

总之, 计算文化心理学致力于应用大数据思维、大数据资源和计算工具来研究文化系统背后的意义制造与互动过程, 它不仅为文化心理分析提供了新的工具, 也为理解人类思维本质和文化形态提供了新的学科范式(Neuman, 2014)。接下来, 我们将主要围绕着"集体主义/个人主义"这一被最广泛讨论的文化心理变量, 介绍计算文化心理学在时间、空间两个维度上的应用: 在时间维度上, 分别从长时程的宏观历史发展和短时程的经济波动转型两个角度介绍计算文化心理学的应用; 在空间维度上, 主要进行生态地理分析, 如文化在特定政治和经济框架内的地区差异, 不同地区生态或户籍制度造成的个人−环境匹配与价值失调。分析框架见图 1。

3　时间维度的文化变迁分析

在时间维度, 研究者们从长时程的宏观历史发展和短时程的经济波动转型切入, 利用谷歌电子图书、谷歌搜索、名字档案、流行歌曲、微博等, 对历史发展、经济波动后的文化变迁进行分析。在方法上, 主要采用了基于关键词词典或特征行为的频次分析、词嵌入联想测验等。

3.1　宏观历史发展与文化心理变迁

文化的形式多种多样, 但几乎所有涉及文化的概念都被构造为个人和群体的关系, 强调个体价值和目标的文化通常被称为个人主义, 而强调群体价值、目标的文化被称为集体主义——个人主义/集体主义也是讨论最为广泛的文化心理变量(Hofstede, 2001)。众多研究表明, 物质财富的增长和社会生态由农村向城市的转变是当今社会最重要的变化趋势之一(陈明星　等, 2009; 李培林, 2002); 进而, 文化价值也发生了巨大变化——世界范围内均出现了个人主义增强、集体主义减弱或受到挑战的趋势(Greenfield, 2009)。

人类文化常常体现在语言当中(Kashima & Kashima, 2003), 而谷歌电子图书恰好提供了一个巨大的、跨度上百年的语料库, 其中语言特征可以用来识别文化价值观及其历史变化趋势, 并可以在 Ngram Viewer 上对目标词或短语进行可视化呈现(Michel et al., 2011)。基于谷歌语料库, 关于个人主义关键词(如自我、独特)和集体主义关键词(如集体、服从)的趋势分析表明, 随着社会现代化与城市化变迁, 英语、汉语中的个人主义词汇使用越来越频繁, 而集体主义词汇使用越来越少或相对稳定(Greenfield, 2013; Zeng & Greenfield, 2015); 以第一人称单数作为个人主义指标的分析表明, 美式英语、汉语、法语、德语、西班牙语、希伯来语、意大利语、俄语在二战以后均表现出个人主义上升的趋势(Twenge et al., 2013; Yu et al., 2016)。当然, 也有研究者基于词嵌入联想测验分析了谷歌中文图书语料库中个人主义/集体主义在不同领域的变迁, 发现在 1950 年至 1999 年间个人主义/集体主义与消极词的关联(向量余弦相似度)大于积极词, 并且个人主义/集体主义与休

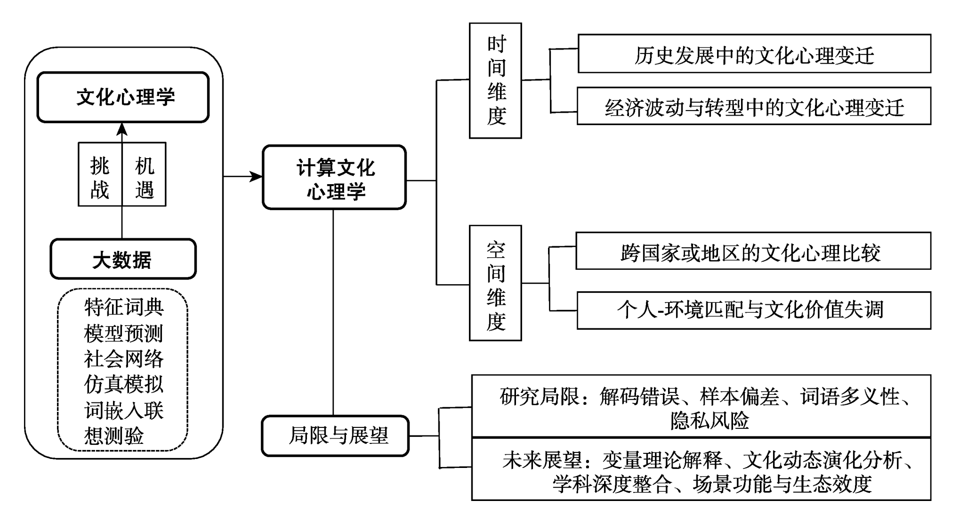


图 1　计算文化心理学的分析框架

闲词的语义关联呈上升趋势，与工作和成就词的语义关联呈下降趋势(Hamamura et al., 2021)；此外，吴胜涛等人(2022)也发现，理性个人主义关键词(如独立、进取)使用频率在 1980 年至 2019 年间呈显著下降趋势，且相对于自我–功利个人主义(如理由、享受)，自我–理性个人主义的语义关联偏好呈显著下降趋势。

在文化实践层面，取名行为、婚姻家庭结构等也是研究文化变迁的重要指标，因为它们反映了一个文化的核心价值观(Lieberson & Bell, 1992)。Twenge 等人(2010)对 1880 年至 2007 年间出生的 3.25 亿美国婴儿名字进行了分析，发现越来越多的美国父母给他们的孩子起独特的名字，希望他们脱颖而出而非迎合大众，这是美国社会个人主义文化越来越盛行的表现。Ogihara (2018)基于离婚率、家庭规模、给宝宝取名和书籍词汇这 4 个指标的实证证据表明，中、日、美三国的文化向更加个人主义的方向转变。

此外，也有研究者从历史发展角度分析文艺作品中的文化心理变化。例如，Dewall 等人(2011)统计了美国流行歌曲歌词中代表个人主义文化的第一人称单数代词和消极情感词使用频率，发现随着城市化、现代化变迁，个人主义呈上升趋势。Cai 等人(2019)研究了中国上世纪 70 年代以来流行歌曲歌词中第一人称代词的变化，也发现了类似现象。可见，相比于传统的文化心理调查，基于语料库、关键词词典和日常生活行为的分析为大尺度的文化趋势分析提供了可能。

总之，基于大型语料库和关键词词典分析，计算文化心理学可精准呈现不同历史发展阶段的文化形态和集体心理，为更好理解大尺度的文化变迁效应提供了新的研究思路。

3.2 经济波动转型与文化心理变迁

文化是人类适应环境的结果，而环境的变化不仅表现为长时程的历史变迁，也表现为短时程的经济波动，后者也会引起人类文化的短期调整。最近的研究表明，在个人主义作为主流价值观且不断增强的美国社会，大萧条时期的青少年却表现出集体主义(如关注他人和环保主义)的增加，并显示出个人主义(如物质主义)减少或趋于平缓的一些迹象(Park, Twenge et al., 2014)。此外，随着经济繁荣或失业率减少时，美国人会更可能给新生儿起不常用的名字，拥护孩子的自主权，并偏爱自我中心的音乐；相反，当经济衰退或失业率增加时，美国人更愿意鼓励孩子帮助他人和被他人喜欢，更加喜欢表达他人取向的音乐(Bianchi, 2016)。随着工业社会各种问题的凸显(如全球变暖、冷战)，千禧一代对“生活目的”的兴趣空前增长，上个世纪 90 年代以来人们对人生目标的兴趣越来越浓厚，个人主义价值观呈平稳下降趋势，集体主义价值观开始复兴(Grant, 2017)。

尽管改革开放以来，中国经历了增长奇迹，但 2010 年以后经济增长放缓，GDP 增长率从 10.6%降至 2012 年的 7.7%，下降了近三个百分点(王少平，杨洋，2017)。为了进一步探究文化作为对生态变化的回应是否会发生短时程的适应性改变，Wu, Li 等人(2018)对 2010 到 2016 年期间的新浪微博活跃用户的行为足迹进行了大数据分析。结果显示，微博用户的集体主义在 7 年中发生了显著变化，其峰值出现在 2011~2012 年的经济衰退期；而在 2013~2016 年经济回到正轨时，集体主义又回到了常态。这表明，中国社交媒体上的文化行为也会随着经济波动而上升或下降(Wu, Li et al., 2018)。上述研究揭示了文化在短期内的演变，即文化对环境的适应性反应并非只有长时程的缓慢渐变过程，也有短时程的即时多变过程。

此外，其他国家或地区的关键政治和经济事件也会对人们的文化心理产生重大影响(Hills et al., 2019)。例如，在俄罗斯经济转型期间，个人主义与集体主义的价值观并非相互排斥，而是彼此共存。在德国纳粹政权时期，不同名字出现在媒体的频率变化可以反映审查制度的存在，某个名字使用频率的降低意味着该名字在此期间受到压制(Alexander, 2017)。

总之，研究者借助大数据技术发现并测算了经济波动、经济转型等短时程的文化变迁效应。通过一些关键的文化心理指标，计算文化心理学捕捉到了常人的生活细节(如婴儿取名)，为重大历史事件(例如经济大萧条、二战)建立了新的注脚。

4 空间维度的文化地理分析

文化地理学认为，文化起源于与特定自然环境相联系的生态系统，是人和自然交互的产物(Anderson, 2020)。具体而言，文化是人类适应环境的结果，不同的环境会产生不同的文化；同时，文化一经产生就具有相对独立性，它在一定条件

下可能契合、也可能背离其成员所处的环境。因此，在地理空间维度，研究者利用 Twitter、Facebook、微博等社交媒体数据和社会调查数据，分析了不同地理空间下各个国家或地区、地域的文化差异，以及文化与环境的契合、背离关系，如“一带一路”沿线的文化多样性，全球范围内不同地区的个人–环境匹配与文化价值失调等。在分析方法上，主要采用基于关键词词典的频次分析、基于机器学习的预测模型以及社会仿真模拟，并将大数据与传统大样本问卷调查方法相结合。

4.1 跨国家或地区的文化心理比较

文化是影响区域合作和经贸往来的重要因素，如个人主义者更愿意与他人建立经贸合作关系，而集体主义者的经贸合作意愿相对较低，并且在交易对象选择上更偏好熟人(Hofstede et al., 2008)。因此，分析“一带一路”沿线国家或地区的民心特点，并找到有效的合作交往模式，是关系到国家战略实施的重大问题。但是，由于地域辽阔、民族众多，且地缘政治、经济、文化因素(如原苏联影响、欧美国家殖民、宗教传统等)异常复杂，传统的抽样统计调查无法做到全样本分析，也很难通过统计调查的方式让作答者准确地进行主观报告。而社交媒体大数据可以提供一个海量、及时的用户行为痕迹的精准数据库，还可以通过建模对某个领域和问题进行预测分析。

吴胜涛等人(2018)以社交媒体 Twitter 单数第一人称代词的使用比例(自我独立性)作为个人主义指标，获得“一带一路”沿线 69 个国家或地区个人主义文化的分布情况；然后，提取世界价值观调查(World Values Survey)数据库中 28 个“一带一路”沿线国家或地区的社会信任度(普遍信任、特殊信任)，并建立个人主义与社会信任的训练模型。最后，以个人主义为自变量，以社会经济学变量为控制变量，建立社会信任的线性回归预测模型。结果表明，“一带一路”沿线国家或地区在个人主义文化指标上存在较大的变异，且主要受欧美国家殖民历史和当地宗教传统的影响，其中欧美国家殖民历史是个人主义最大的影响因素；此外，个人主义较高的国家或地区更依赖陌生人之间的普遍信任，而较少依赖熟人之间的特殊信任；同时，根据结果可将 69 个“一带一路”沿线国家或地区分成三类：欧美模式、伊斯兰模式、前苏联模式。

任孝鹏等人(2017)基于 100 万微博活跃用户，运用词频分析的方法绘制了中国个人主义/集体主义的心理地图。具体而言，采用已经得到验证的中文文本分析软件“文心”系统，利用以往基于个人主义/集体主义的跨文化研究中提到的与个人主义、集体主义有关的词汇，经过讨论后得到个人主义词汇 53 个、集体主义词汇 64 个；然后，根据被试标注的“注册地点”来统计不同省、市、自治区的个人主义/集体主义的词频，比较个人主义和集体主义在省级层面的地区差异。结果发现，南方人的个人主义高于北方人，海外用户的个人主义高于国内用户、集体主义低于国内用户，证明在一个国家内部也存在个人主义/集体主义的地区差异。侯东霞等人(2016)使用国家统计局公布的年度统计数据，发现包括自我雇佣的百分比、离婚率、独居的百分比、65 岁以上老人独居的百分比和三代同堂的百分比等 5 个客观指标的集体主义量表可以预测中国人集体主义的地区差异，即更高的工业化和现代化程度会鼓励个人主义的价值取向。

4.2 个人–环境匹配与文化价值失调

从进化角度来看，人类文化是个体–环境相匹配的结果。例如，在资源匮乏、疾病威胁等相对恶劣的生态环境下，有利于社会绑定的集体主义文化会占据优势，集体主义越高的人主观幸福感也越高(Chiao & Blizinsky, 2010)。然而，个人所习得或持有的文化价值观并非总与环境相匹配。随着人类社会的城市化和人口流动，个人主义在一些地区、人群中会占据主导位置，传统集体主义受到挑战，这就会出现个人–环境匹配困难或文化价值失调——习得或习惯于集体主义文化价值观，而外部生态环境却突然变了，于是容易体会到心理压力；或者，习得了个人主义文化价值观，但外部环境却需要群体互依、合作应对，这时也会出现心理压力或不适应的行为。

考虑到以往小样本调查难以准确回答地区和人口层面这样大尺度的问题，Wu 和 Zhou 等人(2018)采用以词频分析为基础的微博大数据分析和全国大样本调查数据相结合的方法，进行群体和个体两个水平的分析。研究一基于微博基本情绪词表(Weibo-5BML；董颖红 等, 2015)，对 2014 年中国大陆 31 个省/市/自治区的 160 万微博活跃用户的微博内容进行情感分析，结果发现：省级水平的集体主义分数显著地负向预测微博用户的

消极情绪；城市化水平对集体主义与消极情绪的关系起调节作用，即在城市化较低的地区(如贵州、甘肃)，集体主义与微博用户消极情绪显著负相关，在城市化较高的地区(如北京、上海)这一关系则不显著。研究二基于 2010 年中国家庭追踪调查近 3 万人的个体访谈数据，采用多水平回归模型分析，结果发现个体的集体主义显著正向预测积极情绪、负向预测消极情绪；城市化水平对集体主义与主观幸福感的关系起到调节作用，即在城市化水平较低的地区，集体主义越高则积极情绪越多、消极情绪越少；而在城市化水平较高的地区，集体主义对积极情绪、消极情绪的影响均不显著。

面对公共卫生危机或恶劣环境，人类社会数百万年进化而来的集体主义具有显著的生存优势和群体免疫价值，而近代以来在一些国家或地区逐渐发展壮大的个人主义与危机环境不匹配，会导致不适应的行为。新冠疫情期间，Huang 等人(2020)通过新浪微博收集了 2020 年 1 月 11 日到 2 月 21 日期间的集体主义(如我们、合作)和防疫倾向(如洗手、消毒)指数，发现集体主义得分越高，防疫倾向也越强。在集体主义文化盛行的地区，戴口罩是社会责任，是展现作为集体一份子共同抗疫的方式；而在个人主义文化盛行的地区，口罩象征着枷锁和对个人自由的侵犯，不戴口罩彰显个人的选择和自由意志。Lu 等人(2021)通过分析美国所有 3141 个县共计 248, 941 人和美国 50 个州 16, 737 人的口罩佩戴调查数据，以及 29 个国家或地区 367, 109 人和 67 个国家或地区的 277, 219 名 Facebook 用户的调查数据，发现集体主义能显著地预测口罩佩戴率，即集体主义较强的州、国家或地区口罩佩戴率也更高，个人主义越强则口罩佩戴率更低。此外，He 等人(2021)从个人主义/集体主义的生态适应理论出发，分析了中国 1300 万新浪微博用户的数据，发现与气候恶劣的省份相比，温带气候省份的居民更有可能在社交媒体上使用独特的昵称，为个人−环境匹配假说提供了正面案例。

此外，由于研究病毒传播引起的传染病扩散问题不可在真实世界进行实验，因此计算机建模和仿真是研究此类问题的一种有效途径(刘晟，2010)。如黄丽芹等人(2022)通过演化博弈模型和跨文化实验等，发现文化能够预测人们面对 COVID-19 时的防疫成果，即盛行个人主义的国家或地区在疫情防控上效率相对较低(如增加个体违反疫情管控的流动性倾向)，这进一步证实了文化−环境不匹配的社会心理代价。

总之，研究者通过社交媒体的群体分析、全国或全球样本的个体分析以及仿真模拟实验，揭示了不同国家地区或一国内部的文化差异及其心理和行为适应问题。值得注意的是，用传统的大样本调查数据与微博、Facebook 等社交媒体数据相结合的方法重复验证文化心理学领域的个体−环境匹配假说，克服了单纯采用大数据可能导致结果不精确的局限。

5 反思和展望

虽然计算文化心理学已经崭露头角，相对于传统文化心理学研究方法和范式具有无法比拟的优势，但利用大数据进行文化心理分析仍然存在一些局限。

首先，很多研究者忽视了同一变量在不同时代、不同国家的意义分歧，造成解码失真。心理学者检验文化变迁的常用方法是寻找作者认为对文化敏感的词汇并探索其使用频率的时间变化，比如通过追踪第一人称单、复数代词的使用频率来分析集体主义、个人主义的时间变迁效应(Twenge et al., 2013; Yu et al., 2016)。然而，语言总是随着时间演变，同一个词义并不具有稳定性，其不可避免地受到经济、政治或者文化因素的影响。以法语中第一人称单数代词(如 Je, je, J', j', me, moi)为例，它在笛卡尔时代被当作是自我或主体性的指称词来使用，使用频率在 18 世纪末法国大革命期间达到顶峰；但后来被批判为缺乏同一性和确定性的虚幻自我指称，其使用频率也出现下降趋势(Pettit, 2016; Ricoeur, 1990)，因此，检验一个单词随时间的频率变化并不能确保它一定契合研究者的最初意图。作为补救，在选择单词之前，研究者可以通过动态词向量训练来检验单词的词义变迁，以确保所选词的词义符合研究意图(Bamler & Mandt, 2017)。

假定大数据背后的文化成员能代表整个文化，容易将带有偏差的文化成员样本视为研究整体。例如，很多研究者使用当下非常流行的语料库——Google Ngram Viewer 来研究集体主义/个人主义，并把结论推广至整个国家层面(Twenge et al.,

2013)。但实际上，该语料库并不能代表无差别的文化整体，相反，它仅仅是突出了特定子群体的文化，如 Google Ngram Viewer 仅代表书面文化的使用者，社交媒体数据只包含社交媒体用户，而忽视了没有机会接触书籍和社交媒体的文化成员(Pettit, 2016)。所以，研究者要特别注意，诸如图书语料库、社交媒体分析等并不是基于无偏样本，因此在使用该类分析材料时要考虑结论的可推广性和普遍适用性问题。

第三，研究者对量表或题项进行赋值、使得变量可计量，进而将文化心理与行为转换为一种符号系统，但是大数据分析中的频次分析提取的关键词往往具有多义性，如果研究者只是摘取了某个含义 并将其纳入特定理论范畴，则难免会出现选择性谬误。同时，用于处理数据的软件还会进一步固化这一测量误差。例如在 LIWC 软件中，因为“great”属于积极情感范畴，所以“a great amount of rain”这一无关表达将被归类为积极情感表达。这意味着，即便大数据分析技术的飞速发展，但其得出的结果仍不够精确。

第四，存在文化被监控和被操纵、以及侵犯个人隐私的风险。大数据以其海量存储和智能计算优势来提取个人和特定群体的文化和性格数据，进而实施与文化−性格相匹配的精准操纵，甚至攻击。这些做法明显的弊端就是数据隐私性和国家区域安全难以保证，容易被不法分子利用。因此，为了避免个体数据及其分析结果被滥用，全球研究者应该将研究伦理置于大数据分析的首位，共同谋求相应的解决措施和共通的数据法律来解决个人信息及国家或地区文化心理分析的不当使用问题 (Kosinski et al., 2015)。

当然，尽管计算文化心理学存在诸多局限，但我们也不能忽视它在未来的巨大发展空间。首先，从理论高度对大数据测量指标进行准确定义和解释。大数据时代，出现了重视数据驱动、而轻视理论的现象(维克托• 迈尔-舍恩伯格，肯尼斯 • 库克耶，2013)。但实际上大数据只是在数据采集中提供了一个发现或检验科学理论的新方法，不能在知识生产环节中取代理论知识的地位。理论在大数据研究中是重要的，一方面大数据分析的算法模型以及对数据的解释需要理论的支撑(彭理强，2019)，另一方面我们需要运用数据的意向性并由此对数据进行分类，而不同的意向性和不同的分类体系是由理论概念决定的(齐磊磊，2015)。因此，在未来的研究中，研究者的设计和经验数据的使用应在理论指引下进行，尤其在指标维度的构建上要重视理论术语的指称和实体之间的关系(贾向桐，2019)。

其次，利用多种大数据语料库，进行更广泛的文化动态演化分析。过往的研究容易把文化看做静态的单一实体，并且将不同文化之间的异同归因于其背后的国家/民族中根深蒂固的特质，即轻易对群体进行本质化。然而，我们需要意识到文化知识是动态发展的过程，文化的相关潜质或因果连结存在于特定共享文化知识的激活，并且共享文化会随着社会政治和其他类型的社会变化不断地产生和演化(康萤仪 等，2010)。传统的文化心理学研究受制于小样本并容易忽视微弱的相关关系，因此在条件允许的情况下，未来需要依赖多种大数据语料库开展更广泛的纵向大尺度研究，探索文化符号串联及其背后机制的研究线索。例如，研究者可以使用历史档案数据库，来研究共享文化从何而来、其分布与使用的频率如何、共享的边界是什么、共享文化被激活后产生了什么后果等，进而促进共同的和个体的经验。

第三，推动文化心理学与计算机、传播学、历史学等学科的深度整合。文化心理学的力量在于其方法论上的多元化和多学科基础(余霞，钟年，2019; Cohen, 2019)。但是，跨学科也使得学科互补的同时充满文化冲突，例如不同学科会由于不同范式而在考虑问题的优先级等方面存在根本性差异，一旦学科之间产生冲突，问题或许就会贯穿任何主题和学科的边界(Popper, 1963)。然而，当计算思维、数据资源和计算工具被采用后，就可以从方法论层面加速对研究问题的解决。在未来研究中，研究者可以充分利用计算文化心理学固有的学科多样性和数据多源化，尝试通过“计算”去更好地整合文化心理学背后的多个学科视角及其变量维度，使大家共同聚焦某个问题。比如，为了描绘复杂的文化现象共同将目光聚焦于背后多源数据的处理、共享数据库的建成、理论与数据的结合等学科问题，不断推进计算文化心理学的整体发展。

最后，利用大数据的“场景功能”，提升文化心理学研究的生态效度。一直以来，由于实验情境缺乏生态效度，心理学赖以成为“科学”的实验

研究备受质疑；而在大数据背景下，研究者对已有数据的挖掘完全能够在自然情境下生成实验变量，在大数据中挑选情境便是实验方式(喻丰 等, 2015)。在未来研究中，研究者不仅可以聚焦小场景，研究个人经验与近端影响源之间的关系，而且可以探索大场景中更远端影响源与个人经验之间的相互作用，从而分析不同时间、空间的文化差异。值得注意的是，当场景被扩大和拉长，这些真实的场景可能体现了个人行为与动机的绝大部分变异，这体现了自然情境研究的“深度背景化”，有助于提升文化等背景特征的解释水平。

6 总结

总之，计算机和大数据技术深刻影响了文化符号的生产、互动方式以及文化形态本身，并促成了计算文化心理学的产生和发展。通过回顾计算文化心理学的产生过程，以及将大数据分析方法应用到文化变迁、文化地理效应等多个领域的实证研究，有助于我们理解计算文化心理学的学科优势、局限及未来发展方向，并对文化心理研究的思路拓展和方法革新具有启发意义。

尤为重要的是，大数据为大尺度的文化分析提供了可能。文化作为一种宏观且多变的现象，任何一种文化理论都需要从越大规模的时间和空间维度来进行检验和评判，基于大数据的计算文化心理学能够为研究者提供数以亿计的数据以及强有力的分析工具；同时，大数据改变了传统文化心理学的研究范式，更产生了诸多新的文化现象，这蕴含了整个学科思维方式和学科体系的变革(喻丰 等, 2015)。当然，在看到大数据为传统文化心理研究提供新方法、新范式的同时，我们也不能忽视其研究局限，以及潜在的伦理和安全风险。

参考文献

陈浩，乐国安，李萌，董颖红．(2013)．计算社会科学：社会科学与信息科学的共同机遇．*西南大学学报(社会科学版)*, *39*(3), 87–93.

陈明星，陆大道，张华．(2009)．中国城市化水平的综合测度及其动力因子分析．*地理学报*, *64*(4), 387–398.

陈善雄，王小龙，韩旭，刘云，王明贵．(2019)．一种基于深度学习的古彝文识别方法．*浙江大学学报(理学版)*, *46*(3), 261–269.

陈云松．(2015)．大数据中的百年社会学——基于百万书籍的文化影响力研究．*社会学研究*，*30*(1)，23–48, 242–243.

董颖红，陈浩，赖凯声，乐国安．(2015)．微博客基本社会情绪的测量及效度检验．*心理科学*, *38*(5), 1141–1146.

顾肃. (2021)．大数据与认知、思维和决策方式的变革．*厦门大学学报(哲学社会科学版)*, (2), 34–43.

侯东霞，任孝鹏，张凤．(2016)．基于客观指标的中国人集体主义量表．*中国社会心理学评论*, (2), 86–98.

黄丽芹，孙寅，罗思阳．(2022)．个人主义文化价值观对疫情控制效果的影响及其计算心理机制．*心理学报*, *54*(5), 497–533.

贾向桐. (2019)．论当代大数据发展中的理论终结论．*南开学报(哲学社会科学版)*, (2), 124–131.

康萤仪，Michael W. Morris，赵志裕，Veronica Benet-Martínez，张曙光．(2010)．多元文化心灵——文化与认知的动态建构主义进路．*中国社会心理学评论*, (1), 195–216.

李国杰，程学旗．(2012)．大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考．*中国科学院院刊*, *27*(6), 647–657.

李培林. (2002)．巨变：村落的终结——都市里的村庄研究．*中国社会科学*, *23*(1), 168–179+209.

刘晟．(2010)．*面向传染病城市时空传播仿真的个体Agent建模方法研究*(硕士学位论文)．国防科学技术大学，湖南．

刘婷婷，刘箴，柴艳杰，刘翠娟，陈佳鑫．(2016)．一种基于心理学模型的人群踩踏情景仿真方法．*系统仿真学报*, *28*(10), 2448–2454.

罗伯特·怀尔，赵志裕，康萤仪. (2017)．*理解文化：理论、研究与应用/文化心理学精品译丛*．北京：人民出版社．

彭理强．(2019)．大数据主义与经验主义——兼评齐磊磊博士的“大数据经验主义”．*长沙大学学报*, *33*(3), 82–85.

齐磊磊. (2015)．大数据经验主义——如何看待理论、因果与规律．*哲学动态*, (7), 89–95.

任孝鹏，向媛媛，周阳，朱廷劭．(2017)．基于微博大数据的中国人个体主义/集体主义的心理地图．*内蒙古师范大学学报(哲学社会科学版)*, *46*(6), 59–64.

王斌．(2012)．斯图亚特·霍尔的马克思主义语言哲学及其文化研究．*文艺理论与批评*, (2), 40–44.

王少平，杨洋.(2017)．中国经济增长的长期趋势与经济新常态的数量描述．*经济研究*, (06), 46–59.

维克托·迈尔-舍恩伯格，肯尼斯·库克耶. (2013)．*大数据时代：生活、工作与思维的大变革*．浙江：浙江人民出版社．

吴美川．(2020)．马克思恩格斯的语言思想与中国话语建构．*理论探索*, (3), 38–45.

吴胜涛，王予灵，许雅舒，彭凯平．(2022)．更理性，还是更功利？近四十年中国社会两种个人主义的变迁．*审稿中*.

吴胜涛，周阳，傅小兰，刘晓倩，刘天俐，朱廷劭．(2018).

“一带一路”沿线文化与合作交往模式探究：基于社交媒体大数据的心理分析. *中国科学院院刊, 33*(3), 298–307.
吴育锋，刘洪飞，焦冬冬，吴胜涛，朱廷劭. (2018). 小说人物性格的文学智能分析：以《平凡的世界》为例. *中文信息学报, 32*(7), 128–136.
邢付贵，朱廷劭. (2021). 基于大规模语料库的古文词典构建及分词技术研究. *中文信息学报, 35*(7), 41–46.
喻丰，彭凯平，郑先隽. (2015). 大数据背景下的心理学：中国心理学的学科体系重构及特征. *科学通报, 60*(Z1), 520–533.
余霞，钟年. (2019). 文化心理学的跨学科运用——以跨文化传播研究为例. *心理学探新, 39*(5), 393–399.
赵海英，贾耕云，潘志庚. (2016). 文化计算方法与应用综述. *计算机系统应用, 25*(6), 1–8.
Alexander, R. (2017, May). *Identifying and protecting germanic sanctuaries and religious objects from the iron age and the migration period*. Paper presented at the meeting of the 9th european symposium on religious art, restoration & conservation. Iasi, Romania.
Anderson, B. (2020). Cultural geography III: The concept of ‘culture’. *Progress in Human Geography, 44*(3), 608–617.
Bamler, R., & Mandt, S. (2017, July). *Dynamic word embeddings*. Paper presented at the meeting of International conference on Machine learning. Sydney, Australia.
Bianchi, E. C. (2016). American individualism rises and falls with the economy: Cross-temporal evidence that individualism declines when the economy falters. *Journal of Personality and Social Psychology*, *111*(4), 567–584.
Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S. F. (2013, October). *Large-scale visual sentiment ontology and detectors using adjective noun pairs*. Paper presented at the meeting of the Proceedings of the 21st ACM international conference on Multimedia. Barcelona, Spain
Cai, H. J., Huang, Z. H., & Jing, Y. M. (2019). Living in a changing world: The change of culture and psychology. In D. Matsumoto and H. C. Hwang (Eds.), *Oxford handbook of culture and psychology* (pp. 786–817). Oxford, England: Oxford University Press.
Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186.
Chiao, J. Y., & Blizinsky, K. D. (2010). Culture–gene coevolution of individualism–collectivism and the serotonin transporter gene. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 529–537.
Cohen, D. (2019). Methods in cultural psychology. In Kitayama, S., & Cohen, D. (Eds.). *Handbook of cultural psychology* (pp.196–236). New York: Guilford.
de Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *American Institute of Physics*, *1644*(1), 97–104.
de Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review (Glasgow), 65*(3), 122–135.
Desmarchelier, B., & Fang, E. S. (2016). National culture and innovation diffusion. Exploratory insights from agent-based modeling. *Technological Forecasting and Social Change, 105*, 121–128.
Dewall, C. N., Pond, R. S., Campbell Jr., W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts, 5*(3), 200–207.
Favaretto, R. M., Dihl, L., Barreto, R., & Musse, S. R. (2016, September). *Using group behaviors to detect hofstede cultural dimensions.* Paper presented at the meeting of the 2016 IEEE International Conference on Image Processing (ICIP). Arizona, USA.
Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), E3635–E3644.
Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878–1881.
Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129–152.
Grant, G. B. (2017). Exploring the possibility of peak individualism, humanity's existential crisis, and an emerging age of purpose. *Frontiers in Psychology*, *8*, 1478.
Gray, J. (2009). On eScience: A Transformed Scientific Method. In T. Hey, S. Tansley, and K. Tolle (eds.). *The fourth paradigm: Data-intensive scientific discovery* (pp. 19–33). Redmond, WA: Microsoft Research.
Greenfield, P. M. (2009). Linking social change and developmental change: Shifting pathways of human development. *Developmental Psychology*, *45*(2), 401–418.
Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, *24*(9), 1722–1731.
Hamamura, T., Chen, Z., Chan, C. S., Chen, S. X., & Kobayashi, T. (2021). Individualism with Chinese characteristics? Discerning cultural shifts in China using 50 years of printed texts. *American Psychologist, 76*(6), 888–903.
He, L., Chen, Y., & Ren, X. (2021). Climato-economic

origins of variations in uniqueness of nickname on Sina Weibo. *Frontiers in Psychology*, *12*, 599750.

Hills, T. T., Proto, E., Sgroi, D., & Seresinhe, C. I. (2019). Historical analysis of national subjective wellbeing using millions of digitized books. *Nature Human Behaviour*, *3*(12), 1271–1275.

Hofstede, G. J. (2001). *Culture's consequences* (2nd. ed.). Beverly Hills, CA: Sage.

Hofstede, G. J., Jonker, C.M., Verwaart, T. (2008). Individualism and collectivism in trade agents. In: N. T. Nguyen, L. Borzemski, A. Grzech, & M. Ali. (Eds.), *New frontiers in applied artificial intelligence*. IEA/AIE Lecture Notes in Computer Science, 5027. Berlin, Heidelberg: Springer.

Hu, B., Xing, F., Fan, M., & Zhu, T. (2021). Research on the evolution of "Ren" and "Li" in Siku Quanshu Confucian classics. *Frontiers in Psychology*. *12,* 603344.

Huang, F., Ding, H., Liu, Z., Wu, P., Zhu, M., Li, A., & Zhu, T. (2020). How fear and collectivism influence public's preventive intention towards COVID–19 infection: A study based on big data from the social media. *BMC Public Health, 20*(1), 1707*.*

Kashima, Y., & Kashima, E. S. (2003). Individualism, GNP, climate, and pronoun drop: Is individualism determined by affluence and climate, or does language use play a role? *Journal of Cross-Cultural Psychology*, *34*(1), 125–134.

King, A. B. (2001). Affective Dimensions of Internet Culture. *Social Science Computer Review*, *19*(4), 414–430.

Koda, T., & Ishida, T. (2006, January). *Cross-cultural study of avatar expression interpretations*. Paper presented at the meeting of the International Symposium on Applications and the Internet (SAINT'06). IEEE. Arizona, USA.

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543–556.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... van Alstyne, M. (2009). Social science: Computational social science. *Science*, *323*(5915), 721–723.

Lee, I., Choi, B., Kim, J., & Hong, S. J. (2007). Culture-technology fit: Effects of cultural characteristics on the post-adoption beliefs of mobile Internet users. *International Journal of Electronic Commerce*, *11*(4), 11–51.

Lieberson, S., & Bell, E. O. (1992). Children's first names: An empirical study of social taste. *American Journal of Sociology*, *98*(3), 511–554.

Liu, P., Chan, D., Qiu, L., Tov, W., & Tong, V. J. C. (2018). Effects of cultural tightness–looseness and social network density on expression of positive and negative emotions: A large-scale study of impression management by Facebook users. *Personality and Social Psychology Bulletin, 44*(11), 1567–1581.

Lu, J. G., Jin, P., & English, A. S. (2021). Collectivism predicts mask use during covid-19. *Proceedings of the National Academy of Sciences, 118*(23), e2021793118.

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., & Mei, Q. (2016, September). *Learning from the ubiquitous language: An empirical analysis of emoji usage of smartphone users*. Paper presented at the meeting of the Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing. Heidelberg, Germany.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, ... Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Miyamoto, Y., & Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. *Journal of Personality and Social Psychology, 83*(5), 1239–1248.

Neuman, Y. (2014). *Introduction to computational cultural psychology.* Cambridge: Cambridge University Press.

Neuman, Y., Turney, P., & Cohen, Y. (2012). How language enables abstraction: A study in computational cultural psychology. *Integrative Psychological & Behavioral Science*, *46*(2), 129–145.

Ogihara, Y. (2018). Economic shifts and cultural changes in individualism. In A. K. Üskül & S. Oishi (Eds.), *Socio-economic environment and human psychology: Social, ecological, and cultural perspectives*. New York: Oxford University Press.

Park, H., Twenge, J. M., & Greenfield, P. M. (2014). The great recession: Implications for adolescent values and behavior. *Social Psychological and Personality Science*, *5*(3), 310–318.

Park, J., Baek, Y. M., & Cha, M. (2014). Cross-cultural comparison of nonverbal cues in emoticons on Twitter: Evidence from big data analysis. *Journal of Communication*, *64*(2), 333–354.

Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer. *History of Psychology, 19*(2), 141–153.

Popper, F. (1996). Visualization, cultural mediation and dual creativity. *Leonardo, 29*(4), 311–311

Qiu, L., Chan, S. H. M., & Chan, D. (2018). Big data in

social and psychological science: Theoretical and methodological issues. *Journal of Computational Social Science, 1*(1), 59–66.

Ricoeur, P. (1990). Éthique et morale. *Revista portuguesa de filosofía, 46*(1), 5–17.

Tosa, N., Matsuoka, S., Ellis, B., Ueda, H., & Nakatsu, R. (2005, September). *Cultural computing with context-aware application: ZENetic computer.* Paper presented at the meeting of the International Conference on Entertainment Computing. Springer, Berlin, Heidelberg.

Tosa, N., Pang, Y., Yang, Q., & Nakatsu, R. (2019). Pursuit and expression of Japanese beauty using technology. *Arts (Basel), 8*(1), 38.

Twenge, J. M., Abebe, E. M., & Campbell, W. K. (2010). Fitting in or standing out: Trends in American parents' choices for children's names, 1880–2007. *Social Psychological & Personality Science, 1*(1), 19–25.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2013). Changes in pronoun use in American books and the rise of individualism, 1960–2000. *Journal of Cross-Cultural Psychology*, *44*(3), 406–415.

Wu, M. S., Li, B., Zhu, L., Jiao, D., & Zhu, T. (2018, July). *Connecting with others in economic downturn: A big data analysis of collectivism from 2010 through 2016.* Paper presented at the meeting of the Proceedings of The Academic Forum on Research into Social Mentality and Social Psychological Services under the New Era (pp. 81–86). Marietta, Georgia: American Scholars Press, Inc.

Wu, M. S., Zhou, C., Chen, H., Cai, H., & Sundararajan, L. (2018). Cultural value mismatch in urbanizing China: A large-scale analysis of collectivism and happiness based on social media and nationwide survey. *International Journal of Psychology*, *53*(s2), 54–63.

Yu, F., Peng, T., Peng, K., Tang, S., Chen, C. S., Qian, X., ... Chai, F. (2016). Cultural value shifting in pronoun use. *Journal of Cross-Cultural Psychology, 47*(2), 310–316.

Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in china: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, *50*(1), 47–55.

Zhao, Y., Xing, F., Fan, M., Li, H., & Zhu, T. (2021). Psycho-linguistic changes associated with historical celebrities in Henan using classical Chinese big data. *Frontiers in Psychology*, *12*, https://doi:10.3389/fpsyg.2021.648677.

The big data analysis in cultural psychology

WU Michael Shengtao[1], MAO Yunyun[1], WU Shuhan[2], FENG Jianren[3], ZHANG Qingpeng[3], XIE Tian[5], CHEN Hao[5,6], ZHU Tingshao[7]

([1] *School of Sociology and Anthropology, Xiamen University, Xiamen* 361005, *China*) ([2] *School of Journalism and Communication, Xiamen University, Xiamen* 361005, *China*) ([3] *Public Administration School, Guangzhou University, Guangzhou* 510006, *China*) ([4] *School of Philosophy, Wuhan University, Wuhan* 430072, *China*) ([5] *Zhou Enlai School of Government, Nankai University, Tianjin* 300350, *China*) ([6] *Guangzhou Guangdong-Hong Kong-Macao Greater Bay Area Research Center for Social Psychological, Sun Yat-sen University, Guangzhou* 510006, *China*) ([7] *Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, *China*)

Abstract: With the integrated development of big data technology and cultural psychology, computational cultural psychology came into being as a novel interdisciplinary research field, which makes large-scale cultural analysis possible. The key variables of computational cultural psychology are mainly about individualism and collectivism, and the big data technologies (e.g., feature dictionaries, machine learning, social networks analysis, and simulation) have been used to analyze the cultural change effect from the temporal perspective and cultural geography effect from the spatial perspective. It should be noted that there are several limitations in Computational Cultural Psychology, including decoding distortion, sample bias, semasiological variation, and privacy risk, although new method and paradigm are provided. In future directions, theoretical interpretation of variables, cultural dynamics, interdisciplinary integration, and ecological validity should be seriously concerned.

Keywords: big data, culture, computational cultural psychology, cultural change, cultural geography